2019年6月5日 統計数理研究所 オープンハウス

誤特定非正規確率微分方程式モデルの推論

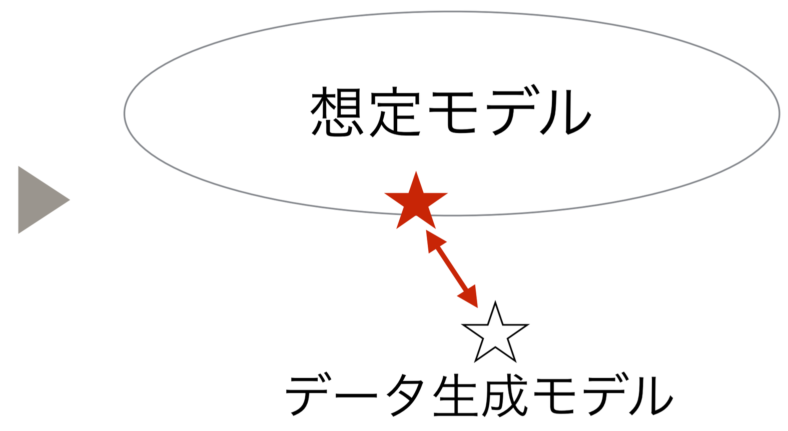
上原 悠槙 社会データ構造化センター 特任助教

モデル設定と誤特定について

データ生成モデル	$dX_t = A(X_t)dt + C(X_{t-})dZ_t$
想定モデル	$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t$

モデル誤特定 (イメージ)

$\forall(\alpha,\gamma)\in\Theta \quad a(x,\alpha)\not\equiv A(x),$
$c(x,\gamma)\not\equiv C(x)$



推定方式

$X_j = X^n_{t_j}, f_j(\theta) := f(X_j,\theta), S(x,\gamma) := c^2(x,\gamma), \phi(\cdot;\mu,\sigma) : dN(0,1)$

1. Drift free estimation of γ :
$$\hat{\gamma}_n := \underset{\gamma\in\Theta_\gamma}{\operatorname{argmax}} \sum_{j=1}^{n} \log\phi\left(X_j; X_{j-1}, h_n S_{j-1}(\gamma)\right)$$
2. Weighted least square estimation α :
$$\hat{\alpha}_n := \underset{\alpha\in\Theta_\alpha}{\operatorname{argmax}} \sum_{j=1}^{n} \log\phi(X_j; X_{j-1} - h_n a_{j-1}(\alpha)), h_n S_{j-1}(\hat{\gamma}_n))$$

- モデル特定下では, $\sqrt{nh_n}$ー一致性, 漸近正規性を持つ: Masuda (2013), Masuda and Uehara (2017)

係数誤特定による影響

{疑似スコア関数} martingale, CLT
$$= \int_0^{nh_n}\int {}^{\exists}h(X_s,z)\tilde{N}(ds,dz) + \frac{1}{\sqrt{nh_n}}\int_0^{nh_n} {}^{\exists}g(X_s)ds$$

- $\int g(x)\pi_0(dx) = 0$, モデル特定下では $g(x)\equiv 0$
- Markov 過程の汎関数時間積分に関する CLT: Bhattacharya (1982), Komorowski and Walczuk (2012), etc…
- 仮定の確認が厄介 & 第1項との同時極限は非自明

拡張 Poisson 方程式 (Kulik and Veretennikov (2013))

We say that a measurable function $f:\mathbb{R}\to\mathbb{R}$ belongs to the domain of the extended generator $\mathcal{A}$ of a homogeneous Feller Markov process X taking values in $\mathbb{R}$ if there exists a measurable function $g:\mathbb{R}\to\mathbb{R}$ such that the process
$$f(X_t) - \int_0^t g(X_s)ds, \qquad t\in\mathbb{R}^+,$$
is well defined and is a local $\mathbb{F}$-martingale with respect to every measure P_x, $x\in\mathbb{R}$. For such a pair (f,g), we write $f\in\mathrm{Dom}(\mathcal{A})$ and $\mathcal{A}f \overset{\mathrm{EPE}}{=} g$.

- 拡散過程 の場合は, 無限小生成作用素に対応する 2次の微分方程式を考えれば良い (Uchida and Yoshida (2011))
- バイアス項の martingale 化 & 表現定理 ⟶ 極限導出

最適値

最適値 $\theta^\star := (\gamma^\star, \alpha^\star)$を
$$\gamma^\star \in \underset{\gamma\in\bar{\Theta}_\gamma}{\operatorname{argmax}}\, \mathbb{G}_1(\gamma)\left(:= -\int_{\mathbb{R}}\left(\log c^2(x,\gamma) + \frac{C^2(x)}{c^2(x,\gamma)}\right)\pi_0(dx)\right),$$
$$\alpha^\star \in \underset{\alpha\in\bar{\Theta}_\alpha}{\operatorname{argmax}}\, \mathbb{G}_2(\alpha)\left(:= -\int_{\mathbb{R}} c(x,\gamma^\star)^{-2}(A(x)-a(x,\alpha))^2\pi_0(dx)\right).$$
で定義する.

- $\gamma^\star$: Stein's loss (log-det divergence) の最小化
- $\alpha^\star$: L2 loss の最小化
- モデル特定下では $\theta^\star$ は真値と一致する

主定理 (Uehara(2018))

適当な正則条件の下,

Theorem I

$${}^{\forall}L>0,\ {}^{\exists}C_L>0,\ \text{s.t.}\ {}^{\forall}r>0,$$
$$\sup_{n\in\mathbb{N}} P(|\sqrt{nh_n}(\hat{\theta}_n - \theta^\star)| > r) \le \frac{C_L}{r^L}$$

Theorem II

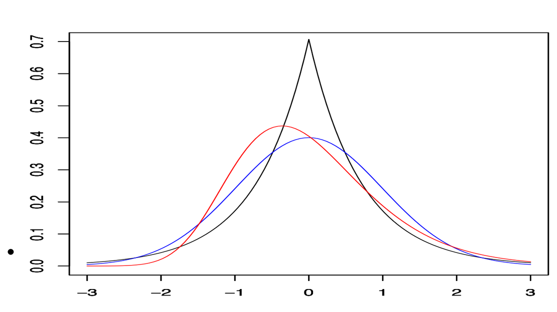
$$\sqrt{nh_n}\begin{pmatrix}\hat{\gamma}_n - \gamma^\star\\ \hat{\alpha}_n - \alpha^\star\end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Gamma^{-1}\Sigma(\Gamma^{-1})^\top)$$

数値実験

データ生成モデル: $dX_t = -\frac{1}{2}X_t dt + dZ_t$

パラメトリックモデル: $dX_t = -\alpha(X_t - 1)dt + \frac{\gamma}{\sqrt{1+X_{t-}^2}}dZ_t$

Z_1 :(i) $\mathrm{NIG}(10,0,10,0)$,
(ii) $\mathrm{bGamma}(1,\sqrt{2},1,\sqrt{2})$,
(iii) $\mathrm{NIG}(25/3, 20/3, 9/5, -12/5)$.



- NIG r.v. : Inverse Gaussian r.v. の正規尺度平均混合
- bGamma r.v. : 二つの独立なGamma r.v.の差

結果

Sample size	frequency	(i) (0.33,1.41)		(ii) (0.37, 1.41)		(iii) (0.37, 1.41)		diffusion (0.33, 1.41)	
n	hn	alpha	gamma	alpha	gamma	alpha	gamma	alpha	gamma
1000	0.05	0.38	1.41	0.40	1.39	0.40	1.39	0.38	1.41
		(0.12)	(0.11)	(0.16)	(0.29)	(0.15)	(0.19)	(0.13)	(0.10)
5000	0.017	0.37	1.41	0.39	1.39	0.38	1.39	0.36	1.41
		(0.09)	(0.08)	(0.11)	(0.23)	(0.11)	(0.15)	(0.09)	(0.08)
10000	0.011	0.36	1.41	0.37	1.39	0.38	1.40	0.36	1.41
		(0.08)	(0.07)	(0.09)	(0.22)	(0.10)	(0.15)	(0.08)	(0.07)